

# Nonlinear softening as a predictive precursor to climate tipping

BY JAN SIEBER<sup>1</sup> AND J. MICHAEL T. THOMPSON<sup>2</sup>

<sup>1</sup>*Department of Mathematics, University of Portsmouth, UK*

<sup>2</sup>*School of Engineering, University of Aberdeen, UK, and Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK*

Approaching a dangerous bifurcation, from which a dynamical system such as the Earth's climate will jump (tip) to a different state, the current stable state lies within a shrinking basin of attraction. Persistence of the state becomes increasingly precarious in the presence of noisy disturbances. We argue that one needs to extract information about the nonlinear features (a “softening”) of the underlying potential from the time series to judge the probability and timing of tipping. This analysis is the logical next step if one has detected a decrease of the linear decay rate. If there is no discernable trend in the linear analysis, nonlinear softening is even more important in showing the proximity to tipping.

After extensive normal form calibration studies, we check two geological time series from paleo-climate tipping events for softening of the underlying well. For the ending of the last ice age, where we find no convincing linear precursor, we identify a statistically significant nonlinear softening towards increasing temperature.

**Keywords:** time series analysis, bifurcation prediction, climate tipping

## 1. Introduction

The on-going work by the United Nations, following the major report of the IPCC [1] in 2007 and the Climate Change Conferences in Copenhagen and Cancun (Mexico), centres on the prediction of future climate change, a key feature of which would be any suggestion of a sudden and (possibly) irreversible abrupt change called a tipping point (Lenton et al. [2], Scheffer [3]). Many tipping points, such as the switching on and off of ice-ages, are well documented in paleo-climate studies over millions of years of the Earth's history.

Predicting such tipplings in advance using time-series data derived from preceding behaviour is now seen as a major challenge, impinging, for example on the possible use of geo-engineering (Launder and Thompson [4]). Techniques introduced by Held and Kleinen [5] and Livina and Lenton [6] draw on the assumption that the tipping events are governed by a bifurcation in an underlying nonlinear dissipative dynamical system. Specifically, researchers search for a slowing down of intrinsic transient responses within the data, which is predicted to occur before most bifurcational instabilities. This is done by determining a so-called propagator, estimated from the correlation between successive elements of a window sliding along the time series (this estimate is called  $\kappa_{ACF}$  in this paper). This propagator, such as the AR(1) mapping coefficient, is a measure of the linear decay rate and should increase to unity at a tipping instability (corresponding to a decrease of the linear decay rate to zero).

Prediction techniques can be tested on climatic computer models, but more challenging is to try to predict real ancient climate tipplings, using their preceding geological data

provided by ice cores, sediments, etc. Using this preceding data alone, the aim would be to see to what extent the actual tipping could have been predicted in advance.

One such study by Livina and Lenton [6] looks at the end of the most recent ice age and the associated Younger Dryas event, about 11,500 years ago, when the Arctic warmed by 7°C in 50 years. This pioneering study used a time series derived from Greenland ice-core paleo-temperature data. A second such study (one of eight made by Dakos et al. [7]), using data from tropical Pacific sediments, gives a good prediction for the end of ‘greenhouse’ Earth about 34 million years ago when the climate tipped from a tropical state into an icehouse state. [Lenton *et al.* this volume] gives a complete overview of current techniques for extracting early-warning signals from time series and compares them on realistic models and a range of paleo-climate time series.

A remark that is made during the analysis of [Lenton *et al.* this volume] is that the absolute values of the extracted quantities have no direct bearing on the tipping probability over time unless one knows also the size of the *basin of attraction*. This basin is determined in the simplest cases by the dominant nonlinear term in the underlying equations of motion, which is the subject of this paper.

## 2. Concepts from Nonlinear Dynamics

The core techniques for analysis of time series of climate data aim to extract the linear decay rate toward (a possibly drifting) noise-perturbed equilibrium [5, 6]. These techniques can be directed at the Earth’s climate as a whole, or at the relevant climate sub-systems described by Lenton et al. [2] as *tipping elements*. Our aim here is to augment this linear information with information about nonlinear features of the underlying dynamics, by examining to what extent we can identify nonlinear softening and include it into the list of early-warning signs for climate tipping.

### (a) *Shrinking basins around dangerous bifurcations*

As we approach a dangerous bifurcation, from which a nonlinear dissipative dynamical system will experience a finite jump to a remote alternative state, the current attractor is located within a shrinking basin of attraction. Correspondingly, the maintenance of the state will become increasingly precarious in the presence of noisy disturbances.

Thinking in terms of an underlying potential energy function (the existence of which is theoretically well-defined for a saddle-node fold and some other bifurcations) the parabolic shape of its graph will become increasingly perturbed by softening nonlinear features.

The relevant bifurcations are the so-called *codimension-one* events (Thompson et al. [9], Thompson and Stewart [10]), namely those that can be typically encountered under a gradual change of a single control parameter. The complete list of the dangerous codimension-one bifurcations is given in Table 1, where we give a brief description of the nature of the shrinking basin. More details can be found in Thompson and Sieber [11]. Focusing first on the local bifurcations (headings (a) and (b) of Table 1) the shrinking boundaries are illustrated schematically in Figure 1, with the control parameter plotted horizontally and the response plotted vertically. In the first picture, Figure 1(a), we illustrate both types of the saddle-node fold bifurcation: the static fold involving a path of equilibrium fixed-points, and the cyclic fold involving a trace of periodic orbits. For these the basin shrinkage is one-sided. The basin is bounded by the unstable side of the parabolically folding path.

<b>(a) Local Saddle-Node Bifurcations</b>	
Static Fold (saddle-node of fixed point)	one-sided basin shrinkage
Cyclic Fold (saddle-node of cycle)	one-sided basin shrinkage
<b>(b) Local Subcritical Bifurcations</b>	
Subcritical Hopf	complete basin shrinkage
Subcritical Neimark-Sacker (secondary Hopf)	complete basin shrinkage
Subcritical Flip (period-doubling)	complete basin shrinkage
<b>(c) Global Bifurcations</b>	
Saddle Connection (homoclinic connection)	outside shrinkage around cycle
Regular-Saddle Catastrophe (boundary crisis)	fractal basin shrinkage
Chaotic-Saddle Catastrophe (boundary crisis)	fractal basin shrinkage

Table 1: Dangerous Bifurcations and the behaviour of the basin of attraction

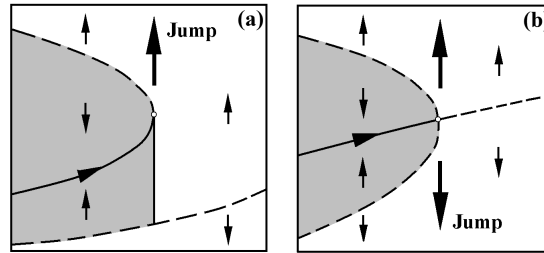


Figure 1: Basin boundary transformations at dangerous bifurcations. (a) the fold bifurcations and (b) the subcritical bifurcations. Solid curves denote stable paths while broken curves denote unstable paths. The basin is shown in grey.

Next, in Figure 1(b), we illustrate the local subcritical bifurcations of which the Hopf, flip and Neimark are codimension-one (we remember that the more familiar static pitchfork bifurcations are codimension-two, being structurally unstable against a symmetry-breaking perturbation). Here we see an unstable solution, which controls the basin boundary, shrinking parabolically around the stable solution, from which the (noise-free) system jumps out of our field of view as the control parameter,  $\mu$ , reaches its critical value of  $\mu_C$ .

The static fold and the Hopf bifurcation have a theoretically well-defined potential energy surface that neatly summarizes the shrinking basin as illustrated in Figure 2. Note that Figure 2(a) will be discussed more fully in section 2(b).

#### (b) A closer look at the fold

The fold is the simplest and most common way in which an equilibrium of a nonlinear dissipative dynamical system can lose its stability. Having only a single active degree of freedom ( $x$ ) and being observable under the variation of a single control parameter ( $\mu$ ) it can be illustrated on a graph of  $x$  against  $\mu$  as a smooth curve that simply reaches an extreme value (a maximum, say) of the control  $\mu$ , as shown in Figure 1(a). So although it is traditionally called a saddle-node bifurcation, it does not exhibit any obvious ‘bifurcation’ of a path, but rather just a smooth folding of an existing path. In the case of a static fold (to which we shall largely restrict our attention) the path is a trace of equilibrium fixed points, while for a cyclic fold the path would be a trace of periodic solutions.

Assuming that we are close to a fold, the intrinsic damping of the system will have become super-critical so the system will have the non-oscillatory response of an over-damped

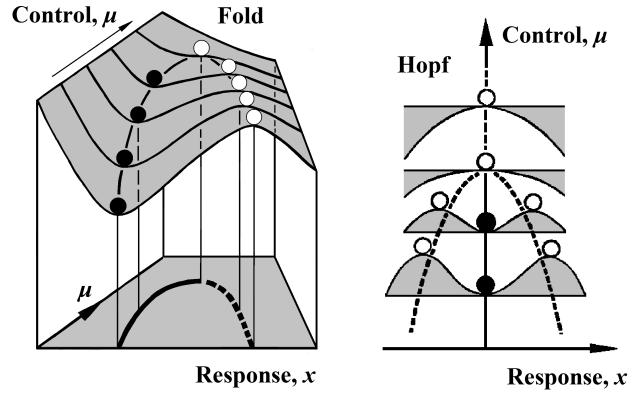


Figure 2: Total potential energy transformations in (a) the saddle-node fold and (b) the Hopf bifurcation (via averaging). Black balls denote stable equilibrium states while white balls denote unstable states.

particle sliding (in thick oil, as we might imagine) on a notional potential energy surface  $U(x, \mu)$ . This is illustrated for the fold in Figure 2(a), where we show the potential energy surface erected over the  $(x, \mu)$  base plane. Here the solid part of the equilibrium path, corresponding to energy minima, is a stable node, while the dashed part, corresponding to energy maxima (more generally a geometrical saddle in higher dimensions) is a dynamically unstable saddle.

We see that as the control  $\mu$  is slowly increased towards its critical value,  $\mu_C$ , the stable equilibrium solution gets closer and closer to the hill-top potential barrier, and so gets progressively more precarious in the presence of noisy disturbances. With  $\mu$  increasing at an infinitesimal rate, noise will ensure escape over the hill-top before  $\mu$  reaches  $\mu_C$ . If, however,  $\mu$  increases rapidly, this early escape may be delayed or suppressed as we have examined quantitatively elsewhere (Thompson and Sieber [12]).

To give a greater perspective to our view of the fold it is useful to introduce another control parameter  $\mu_2$  such that our parameter space is now represented jointly by the  $(\mu_1, \mu_2)$ -plane. We can now erect an equilibrium surface, of height  $x$ , over the two-dimensional control plane as illustrated in Figure 3. Here we see two fold lines whose projections divide the control plane into regimes exhibiting either one or three co-existing solutions. These fold lines coalesce and vanish at the cusp. This cusp is a codimension-two phenomenon, which is typically only observed when we have independent control of both parameters ( $\mu_1$  and  $\mu_2$ ). This is made transparent by the fact that a typical path in control space cannot be expected to pass precisely under the cusp point; by contrast, if a parameter path passes through a fold line all near-by paths also cross this fold line, which is what ‘the fold has codimension-one’ means. For an appropriate choice of path in the  $(\mu_1, \mu_2)$ -plane through the cusp one sees a pitchfork bifurcation (Thompson and Stewart [10]).

We will consider the case in which the main equilibrium sheet (in grey) is stable while the narrow inverted sheet (white) is unstable.

A climate system will have a large number of possible control parameters, but in trying to predict a tipping point we will be studying a particular evolution which corresponds to a particular route in control space; this is precisely why we are only likely to encounter a fold, and not a cusp in Figure 3. To explore possible scenarios that might underlie a tipping study, it is now useful to examine some possible routes in our two-parameter model of Figure 3.

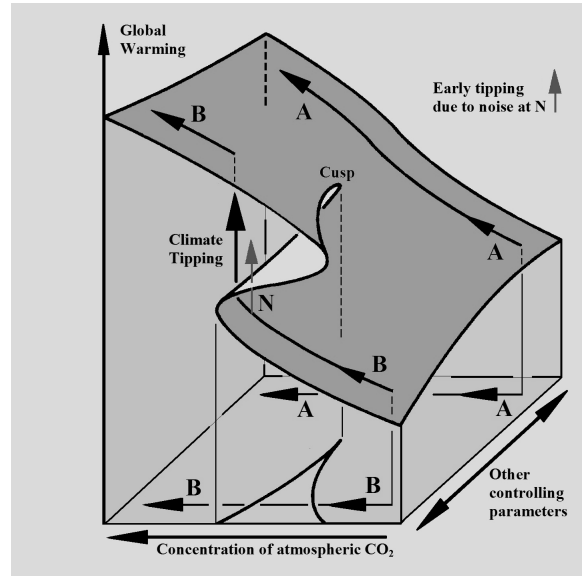


Figure 3: The cusp catastrophe and its associated pattern of folds. The notional axes show how such a cusp might arise in a climate tipping scenario.

Consider, first, the simple scenario in which the system parameters take path  $\overline{BB}$  leading transversally through a fold line. This is the classical encounter with a fold that is implicitly envisaged in earlier work. In the presence of noise a premature tip from N occurs with a certain probability depending on the noise level and the speed with which the path is traversed, as analysed by Thompson and Sieber [11]. If the noise level were particularly high, a jump from N could easily be perceived as a purely noise-induced jump with no bifurcation, even though it is actually caused by the proximity to the fold: this could hold even if there were no movement in the control space at all, the system having rested at N throughout ([Ashwin *et al.* this volume] call this N-tipping). Another scenario, not specifically illustrated in the figure, could arise if the route in control space approached the fold line, but then turned back away from it. Analysis would then show a temporary decrease of the strength of the attraction (with the AR(1) coefficient moving towards unity) which might be discounted as a false alarm, even though the system did approach a fold and a noise-induced escape was temporarily probable.

Finally, in the scenario of path  $\overline{AA}$ , one would observe a temporary increase of the AR(1) coefficient, even though no bifurcation is apparent but rather a gradual shift of the steady state. This shows that a rapid transition can be related to a nearby fold that was just barely missed. Similarly, the scenario called R-tipping by [Ashwin *et al.* this volume] occurs if one changes the other control parameter in Figure 3 sufficiently rapidly from the path  $\overline{AA}$  to N (avoiding the folds). In this scenario the system jumps to the other stable equilibrium despite having never crossed a bifurcation curve.

### 3. Time series analysis for leading order terms — qualitative overview

The extraction of the underlying dynamics from time series is divided into subtasks of increasing difficulty. Assume that we have a time series coming from a process that is either stationary or has a slowly drifting underlying parameter  $\mu$  as suggested by the paths in

Figure 3. In order to predict (or identify) the dynamics one attempts to extract the leading orders of the dominant components of the underlying equations of motion. We illustrate

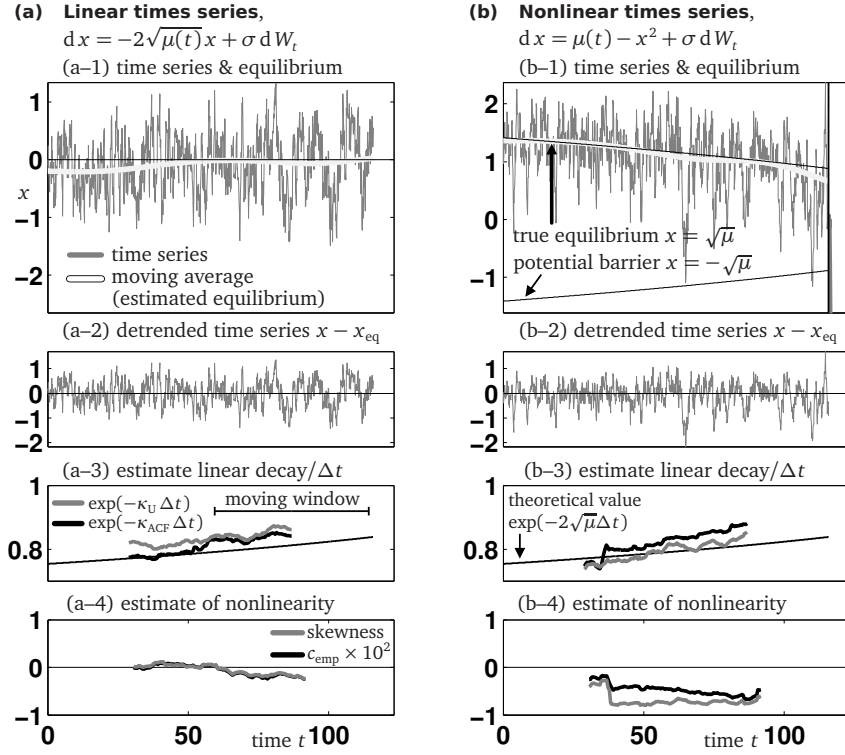


Figure 4: Comparison of time series generated by linear (left column, (a-1) to (a-4)) and nonlinear process (right column, (b-1) to (b-4)). Both time series are stationary and have identical linear decay rates. As measures for nonlinearity we used skewness and the coefficient  $c_{\text{emp}}$  (as defined below in equation (3.5)). Parameters:  $\mu(0) = 2$ ,  $N = 1223$ ,  $\Delta t = 0.1$ ,  $\sigma = 1$ , window length for linear and nonlinear analysis,  $w = N/2$ ,  $\varepsilon = 0.01$ .

the steps using a time series obtained from a process which has the saddle-node normal form with a slowly drifting normal form parameter  $\mu$  as its deterministic part (corresponding to Figure 2(a)) and is influenced by additive Gaussian noise. We permit the parameter  $\mu$  to drift slowly with speed  $\varepsilon$ :

$$\begin{aligned} dx &= [\mu - x^2] dt + \sigma dW_t, \text{ where} \\ d\mu &= -\varepsilon dt. \end{aligned} \quad (3.1)$$

Here  $x$  is the response,  $\mu$  is the control parameter, and  $t$  is the time. The Gaussian perturbation  $dW_t$  has zero mean and unit variance such that  $\sigma$  controls the noise induced variance added to the dynamics of the deterministic saddle-node form  $\dot{x} = \mu - x^2$ , which has the stable equilibrium  $x_{\text{eq}} = \sqrt{\mu}$ . The second equation prescribes the linear sweep of  $\mu$  from its starting value  $\mu_0 > 0$  towards zero at rate  $\varepsilon \ll 1$ . For fixed  $\mu > 0$  the unperturbed system ( $\sigma = 0$ ) has a stable equilibrium at  $x = x_{\text{eq}} = \sqrt{\mu}$  and an unstable equilibrium at  $x = -\sqrt{\mu}$ .

In Section 4 we will use the saddle-node normal form (3.1) to test the ability of a range of estimators to distinguish a time series generated by a process with a quadratic nonlinearity from a linear time series. Two example time series and their analysis are shown in

Figure 4. Both time series have identical linear properties but differ in the dominant non-linear term of the deterministic part of the dynamics. The softening of the full (estimated) nonlinear potential well (corresponding to the illustration in Figure 2(a)) is displayed in Figure 5.

(a) *Equilibrium — zero order*

The location of the slowly drifting equilibrium and its trend is estimated first. [Lenton *et al.* this volume] call this step detrending and propose either piecewise linear fitting or filtering with a Gaussian kernel. The result of filtering with a Gaussian kernel is shown as a thick grey curve in the row 1 of Figure 4. Every method for detrending requires a bandwidth parameter (for example the width of the Gaussian kernel). For Figure 4 we chose the bandwidth which the kernel density estimation routine (Matlab's `kde` by Botev *et al.* [14]) offered. The thin black curve in Figure 4(b-1) shows the true value of the quasi-equilibrium (for  $\varepsilon = 0$ ). It gives a visual estimate of how the reliability of the zero-order estimates depends on the quality of the data. The oscillatory deviations of the extracted equilibrium (thick light grey curve) from its true value (thin black curve) shows that the automatically chosen bandwidth was slightly smaller than the theoretical optimum.

(b) *Linear decay rate — first order*

Assuming that the deterministic part of the underlying process has a stable equilibrium  $x_{\text{eq}}$  (which is possibly slowly drifting), and that the zero-order estimate is an approximation for  $x_{\text{eq}}$ , the next step is to estimate the linear decay rate  $\kappa$  toward  $x_{\text{eq}}$  and the trend of  $\kappa$  over time. Generally, if the underlying deterministic process is high-dimensional then these estimates capture the dominant (that is, smallest) decay rate. They are typically applied to the detrended time series, that is,  $\tilde{x} = x - x_{\text{eq}}$ , which is shown in the second row of Figure 4.

- **ACF** The  $k$ -step (usually one-step) autocorrelation function (ACF)  $\alpha$  is fitted directly to the detrended time series:  $\tilde{x}_{k+1} = \alpha \tilde{x}_k$ . This was introduced by Held and Kleinen [5] as degenerate finger-printing for the analysis of climate time series. The ACF  $\alpha$  is related to  $\kappa$  via

$$\alpha = \exp(-\kappa_{\text{ACF}} \Delta t).$$

The estimate for  $\kappa_{\text{ACF}}$  is shown in Figure 4(a-3) and (b-3).

- **DFA** Detrended fluctuation analysis has been introduced by Livina and Lenton [6] for climate time series, see also [Lenton *et al.* this volume] for a short description of how one extracts  $\kappa$  from this analysis.
- **Quasi-stationary density** If one assumes that the deterministic dynamics of the detrended time series  $\tilde{x}$  is essentially that of an overdamped particle moving in a potential well  $U(\tilde{x})$ , that is,

$$d\tilde{x} = -\partial_x U(\tilde{x}) dt + \sigma dW_t, \quad (3.2)$$

then the linear decay rate equals the second derivative  $\partial_{xx} U(0)$  of  $U$  at the bottom of the well,  $\tilde{x} = 0$ . Moreover, the stationary density  $p(x)$  is related to the shape of the

potential well  $U$  via the stationary Fokker-Planck equation

$$\begin{aligned} \frac{1}{2} \partial_{xx} p(\tilde{x}) &= \partial_x [-\sigma^{-2} \partial_x U(\tilde{x}) p(\tilde{x})], \text{ which gives, integrated once,} \\ \frac{1}{2} \partial_x p(\tilde{x}) &= -\sigma^{-2} \partial_x U(\tilde{x}) p(\tilde{x}) + \sigma^{-2} c, \end{aligned} \quad (3.3)$$

where  $c$  is the constant of integration. This constant satisfies  $c = \lim_{\tilde{x} \rightarrow \pm\infty} \partial_x U(\tilde{x}) p(\tilde{x})$ , and can be interpreted as the flow rate ( $c < 0$  indicates flow toward  $-\infty$ ). Remembering that the drift speed  $\varepsilon$  of the system parameter  $\mu$  is small, the true time-dependent probability density  $p$  still satisfies (3.3) approximately at each  $\mu$ :

$$\frac{1}{2} \partial_x p(\tilde{x}, \mu) = -\sigma^{-2} \partial_x U(\tilde{x}, \mu) p(\tilde{x}, \mu) + \sigma^{-2} c(\mu) + O(\varepsilon). \quad (3.4)$$

In (3.4) we included the dependence of all quantities on  $\mu$  expressly and remember that  $\dot{\mu}(t)$  is of order  $\varepsilon$ . Assuming quasi-stationarity, we neglect the order- $\varepsilon$  terms and determine the coefficients,  $-\sigma^{-2} \partial_x U$  and  $\sigma^{-2} c$ , by fitting the empirical density  $p_{\text{emp}}(x)$  from a time window  $[t - w/2, t + w/2]$  to the Equation (3.4). Specifically,  $c_{\text{emp}} = \sigma^{-2} c$  is a scalar and  $\partial_x U(0)$  is equal to 0 after detrending such that one has to fit two coefficients,  $\kappa_U$  and  $c_{\text{emp}}$  using (3.4) if one truncates  $\partial_x U$  after first order:

$$\frac{1}{2} \partial_x p_{\text{emp}}(\tilde{x}) = -\kappa_U \tilde{x} p_{\text{emp}}(\tilde{x}) + c_{\text{emp}}. \quad (3.5)$$

In problems where no escape is possible one can drop the term  $c$  in (3.3) (and, thus, in (3.4) and (3.5)), and simplify (3.3) to

$$U = -\frac{\sigma^2}{2} \log p. \quad (3.6)$$

Livina et al. [15] used relation (3.6) (fitting of  $U$  with higher-order polynomials) to detect potential wells in time series that included rapid transitions. Fitting  $U$  with a parabola (where  $U(0)$  is irrelevant and  $\partial_x U(0) = 0$ ) is equivalent to using the variance  $\sigma_{\text{emp}}^2$  of  $p_{\text{emp}}(\tilde{x})$  if the density  $p_{\text{emp}}$  is Gaussian. Ditlevsen and Johnsen [16] state that monitoring the empirical variance  $\sigma_{\text{emp}}^2$  helps to avoid erroneous detection of false trends.

[Lenton *et al.* this volume] compare these three estimates,  $\kappa_{\text{ACF}}$ , the DFA-based estimate for  $\kappa$ , and the variance  $\sigma_{\text{emp}}^2$  (which is inversely proportional to  $\kappa_U$ ), in detail using time series arising in climate models and in palaeoclimate records. Methods based on properties of the spectrum of the time series were proposed and investigated by Kleinen et al. [17], Biggs et al. [18] but are not discussed here. Figure 4 shows in row 3 how the estimates of  $\kappa_U$  from the quasi-stationary density compare to the ACF estimates  $\kappa_{\text{ACF}}$ . For both time series the estimates are quantitatively close to the theoretical value of  $\kappa$  (which is indicated by the thin black line in Figure 4(a-3,b-3)). We also observe that in Figure 4(a-3,b-3) the local trends of  $\kappa_{\text{ACF}}$  from the autocorrelation estimate and of  $\kappa_U$  based on the quasi-stationary density are strongly correlated, so it is unclear if monitoring both quantities really prevents false alarms.

There is one notable difference between the AR(1) and the DFA estimate on one hand and the distribution-based estimate  $\kappa_U$  on the other hand: the estimate for  $\kappa_U$  based on the



quasi-stationary density only looks at the distribution of the values  $\tilde{x}_k$  in the window of interest but does not care about the order in which they appear. This is in contrast to, for example, the AR(1) estimate which determines the correlation between each value  $\tilde{x}_k$  and its predecessor.

#### *Estimate of the input noise amplitude*

If the AR analysis shows the presence of a single positive dominant coefficient (which gives evidence that the underlying dynamics has really one distinct direction in which the decay is slowest) then one can combine both estimates of the decay rate to obtain an estimate of the amplitude  $\sigma$  of the additive noise. Equations (3.3) and (3.5) imply that  $\kappa_U$  is an estimate for  $\sigma^{-2}\kappa$  where  $\kappa$  and  $\sigma^2$  are the true linear decay rate and input noise amplitude. Thus, if we replace the unknown  $\kappa$  by the estimate  $\kappa_{\text{ACF}}$  we can use the relation between  $\kappa_U$  and the true  $\kappa$  to estimate  $\sigma^2$ :

$$\sigma_{\text{emp}}^2 = \kappa_{\text{ACF}} / \kappa_U \quad (3.7)$$

where  $\kappa_{\text{ACF}}$  is the estimate of the linear decay rate obtained from the autocorrelation and  $\kappa_U$  is the estimate obtained from the quasi-stationary density  $p_{\text{emp}}$ . This is an alternative to using the residual of the least-squares fit that one obtains when estimating  $\kappa_{\text{ACF}}$ . We found that the residual systematically underestimates the noise level in the normal form examples shown in Fig. 4.

#### *(c) Nonlinearity — basin boundary*

One problem left open by the linear analysis is that the quantitative value of the estimated linear decay rate  $\kappa$  and the estimated noise-level do not give any indication for the probability of tipping. What the linear methods estimate is the  $\alpha$  and the  $\sigma$  of a discrete process

$$\tilde{x}_{k+1} = \alpha \tilde{x}_k + \sigma \eta_k \quad (3.8)$$

(where the  $\eta_k$  are assumed to be independent and normally distributed random numbers). The discrete process (3.8) is the linearisation of the time- $\Delta t$  map of the continuous process (3.2). An estimate for  $\alpha$  less than but close to unity does not necessarily indicate that we are close to a stability boundary for the nonlinear problem (3.2). Rather, it implies that the time step  $\Delta t$  between successive measurements has been small compared to the mean decay time to half,  $\log 2 / \kappa$ , of the process:

$$\alpha(t_k) = \exp(-\kappa(t_k)\Delta t) = 1 - \kappa(t_k)\Delta t + O((\Delta t)^2).$$

The *trend* of the estimated  $\alpha$  is an indicator for incipient tipping because (if extrapolated in any way) it gives an estimate for the time to tipping if one ignores the possibility of early escape. However, the trend of  $\alpha$  is less certain even in the artificial time series of Figure 4. Similarly, the estimated noise amplitude  $\sigma$  has to be compared to the coefficient in front of the leading nonlinear term of the right-hand side (which equals  $-1$  in (3.1)). The ratio between  $\sigma$  and nonlinear term enters the estimates for the probability of tipping if there is no discernible upward trend in  $\alpha$  (N-tipping), or for the probability of early escape if the trend of  $\alpha$  points upwards (see Thompson and Sieber [12]). So without at least an order-of-magnitude estimate of the leading nonlinear term the linear methods only provide an estimate for the tipping time if there is a clearly discernible trend in  $\alpha$ , and even then this estimate discounts early escape.

Figure 4 illustrates this problem. Apart from a slow trend both time series in Figure 4 have identical properties at the linear level. However, while for the time series in Figure 4(b-1) the median time to tipping is  $t = 100$  according to Thompson and Sieber [12] (indeed, it escapes shortly after  $t = 120$ ), for the time series in Figure 4(a-1) the extrapolation (correctly) predicts tipping (or rather, linear instability) at  $t \approx 320$ . Thus, we observe that the linear properties of the time series alone are not sufficient to provide estimates about tipping, for example, the median time to tipping. The difference between the time series in Figure 4(a-1) and Figure 4(b-1) is in the dominant nonlinear term of the deterministic part. That this term plays a role is not surprising because tipping is a feature of nonlinear systems. Extracting its presence from a time series is a necessary step that has to follow the linear analysis. The approaches presented in the following all generalise

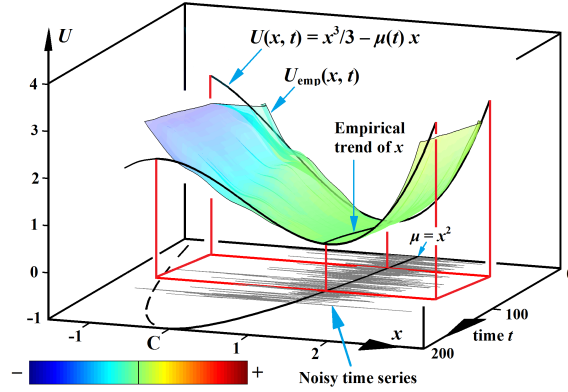


Figure 5: Nonlinear potential energy surfaces extracted from noisy time series using sliding windows ( $w = N/2$ ) for the saddle node fold ( $U_{\text{emp}} = -\sigma_{\text{emp}}^2 \log(p_{\text{emp}})/2$  where  $\sigma_{\text{emp}}$  was estimated using (3.7)). At either end of the surface we have plotted the known function  $U(x, t) = x^3/3 - \mu(t)x$  ( $\sigma = 1$ ,  $\varepsilon = 0.01$  and  $\mu(0) = 2$ ). Red denotes a positive deviation, blue a negative deviation from the parabola given by linear theory.

the estimate for  $\kappa_U$ , based on the quasi-stationary density, by looking at the full nonlinear potential well  $U_{\text{emp}}(\tilde{x})$  associated to the quasi-stationary density  $p_{\text{emp}}(\tilde{x})$  via the Fokker-Planck equation (3.4). Figure 5 illustrates how  $U_{\text{emp}} = -\sigma^2 \log(p_{\text{emp}})/2$  looks like for the time series shown in Figure 4 column (b). The realisation of the noisy, evolving time series is shown in the base plane together with the equilibrium path  $\mu = x^2$ , and the estimated (empirical) potential,  $U_{\text{emp}}$  is shown as the coloured surface. The estimate for the potential well is taken in a sliding window, which explains why the surface can only be determined in the central region of the time series with half the length of the window unrepresented at either end. The numerically estimated equilibrium trend  $x_{\text{eq}}$  is displayed as a black curve at the bottom of the valley of  $U_{\text{emp}}$ . For comparison with this  $U_{\text{emp}}$  we display at either end the real (nonlinear) potential  $U(x, t)$  corresponding to (3.1), which at the later time is already showing the hill-top to the left. The colour in Figure 5 shows the deviation of the empirical potential from the linear-theory parabola, which is not itself displayed. The softening, highlighted by the blue colouration, is seen on the left hand side which is approaching the hill-top. Transparency of the colour is used to show the number of data points that support a particular part of the surface  $U_{\text{emp}}(\tilde{x})$  (as given by the empirical quasi-stationary density  $p_{\text{emp}}(\tilde{x})$ ).

Several methods have been proposed to detect nonlinearity in the potential well in the form of this type of softening (or tilting).

- **Skewness** Guttal and Jayaprakash [19, 20] studied how the closeness of a bifurcation influences the skewness  $\gamma$  of the empirical quasi-stationary density from time series. As this approach is also based on the empirical quasi-stationary density  $p_{\text{emp}}$ , it generalises the estimate  $\kappa_U$  to non-parabolic features of the well  $U$  and non-Gaussian features of the stationary density. Guttal and Jayaprakash [19, 20] proposed to look at the *trend* of the skewness  $\gamma$  over time to detect incipient bifurcations. Figure 4(b-4) shows how the skewness, taken over a single window, changes. It is noticeably negative (compare to the empirical skewness from the linear time series in Fig 4(a-4)) but no trend is discernable.
- **Quasi-stationary density** Livina et al. [15] generalised the potential well analysis for the quasi-stationary density beyond the estimate  $\kappa_U$  for the linear decay rate, fitting the potential well to higher-order polynomials of even degree using relation (3.6). This analysis was not used to attempt prediction of transition probabilities from time series but it was applied to time series that already included all transitions to count the number of wells of  $U$  depending on time  $t$  (which corresponds to the number of modes (peaks) of the empirical quasi-stationary density  $p_{\text{emp}}$ ). See also [Beaulieu *et al.* this volume] for methods that detect change points in time series.
- **Drift ratio** If one assumes that the underlying system parameter is approaching a fold more or less linearly ( $\dot{\mu}(t) = \varepsilon > 0$ ) and one is sufficiently close to the fold already then the ratio between the drift  $\Delta x_{\text{eq}}$  of the equilibrium estimate and the increase of the estimated linear decay rate  $\Delta \kappa$  estimates the quadratic term in the normal form [12].

#### 4. Quantitative estimates of the nonlinear coefficients

The colour shading of Figure 5 suggests that the nonlinear term in the underlying deterministic system shows up in the quasi-stationary density  $p_{\text{emp}}$ . This raises two questions. First, can the level of deviation from linear theory be distinguished with confidence from random chance? That is, can one state, when looking at a time series, that a significant quadratic term is present? Second, is it possible to quantify the size of the nonlinear term with reasonable level of certainty?

Finding a significant nonlinearity is far easier than estimating its precise value because biases introduced during the analysis of the equilibrium and decay rate (zero-order and first-order analysis) may distort the value of the nonlinear estimate but may still keep it significantly different from zero. We will demonstrate this in Section 5 for two paleo-climate time series.

##### (a) Properties of the stationary density of the saddle-node normal form

Figure 6 shows the dependence of the mean, the variance  $\text{var } p$ , and the skewness  $\gamma$  of the empirical stationary density  $p_{\text{emp}}$  on the system parameter  $\mu$  for (3.1) as computed using the stationary Fokker-Planck equation (3.3). Note that this corresponds to the stationary process where the random realisation is re-injected at  $+\infty$  whenever it has escaped to  $-\infty$ . We observe that the dependence of  $\gamma$  is strongly nonlinear for the saddle-node normal form

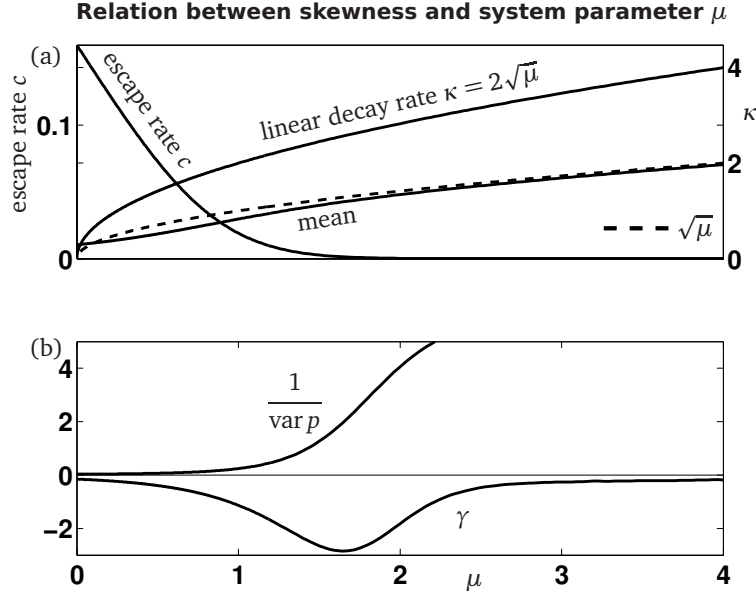


Figure 6: Dependence of mean, variance  $\text{var } p$  and skewness  $\gamma$  of the empirical stationary density for the saddle-node normal form with additive Gaussian noise (3.1) as computed directly from the Fokker-Planck equation (3.3) with  $-\partial_x U(x) = \mu - x^2$  and  $\sigma = 1$ . The escape rate  $c$  is defined by the constant of integration in Equation (3.3).

with noise ( $\dot{\mu}(t) = \varepsilon = 0$ ,  $\sigma = 1$  in Equation (3.1)). This means that trends of the skewness are likely to be difficult to ascertain when approaching a fold. However, the presence of skewness is an indicator for the nonlinearity uniformly for all parameters  $\mu$ , and its sign indicates the direction of escape.

Note also how the nonlinearity affects the empirical mean and the variance  $\text{var } p$ , which is no longer inversely proportional to the linear decay rate  $\kappa$ . This shift of the mean and the additional variance is in part due to the fraction of trajectories that escape, in part due to the non-parabolic shape of the well at some distance from its local minimum.

(b) *Practical estimates of the nonlinear term from time series*

We can estimate the nonlinearity directly as the deviation of the empirical potential well  $U$  from a parabola as suggested by figure 5 (we are only interested in  $\partial_x U$ ). The simplest approach (apart from looking at skewness  $\gamma$ ) is to fit  $\kappa_U$  and  $c_{\text{emp}}$  to the empirical density  $p_{\text{emp}}$  using Equation (3.5). If  $\partial_x U(\tilde{x})$  is indeed linear then  $c_{\text{emp}}$  would be zero such that  $c_{\text{emp}}$  is a signed scalar measure for the deviation of  $\partial_x U(\tilde{x})$  from linearity, serving as a proxy for the present nonlinearity. A second alternative is to expand  $-\partial_x U(\tilde{x})$  to second order, incorporating a quadratic nonlinearity directly into  $\partial_x U$  with an unknown coefficient  $N_2$ . This leads to

$$\begin{aligned}
 -\sigma^{-2} \partial_x U(\tilde{x}) &= -\kappa_U \tilde{x} + N_2 \tilde{x}^2, \quad \text{such that we fit} \\
 \frac{1}{2} \partial_x p_{\text{emp}}(\tilde{x}) &= [-\kappa_U \tilde{x} + N_2 \tilde{x}^2] p_{\text{emp}}(\tilde{x}) + c_{\text{emp},2}.
 \end{aligned} \tag{4.1}$$

Both quantities,  $c_{\text{emp}}$  and  $N_2$ , measure the deviation of  $-\sigma^{-2} \partial_x U(\tilde{x})$  from its linear approximation  $\kappa_U \tilde{x}$ . The main difference between them is the weighting ( $\tilde{x}^2$  for  $N_2$ , uniform for  $c_{\text{emp}}$ ). In summary, the step-by-step procedure to extract an estimate of the quadratic term (or a proxy for it) from a time series  $x_k$  is the following.

1. Detrend the time series  $x_k$  (see Section 3(a)). We call the detrended time series  $\tilde{x}_k$ . The following steps are performed for each suitable time  $t$  to obtain  $\kappa_U(t)$ ,  $c_{\text{emp}}(t)$ ,  $N_2(t)$  or skewness  $\gamma(t)$ :
2. Choose a length  $w$  of sliding windows and obtain the empirical stationary density  $p_{\text{emp}}(\tilde{x})$  for the sliding window centered at  $t$ . (We used `kde1d` to obtain  $p_{\text{emp}}(\tilde{x})$ . For non-stationary time series we used the sliding window length  $w = N/2$ .)
3. Use relation (3.5) to find  $\kappa_U(t)$  and  $c_{\text{emp}}(t)$  from  $p_{\text{emp}}$ . Use relation (4.1) to find  $N_2(t)$  (and another estimate for  $\kappa_U$ ) from  $p_{\text{emp}}$ . Compute the empirical skewness  $\gamma(t)$  from  $p_{\text{emp}}$ .

(c) Comparison of estimates using saddle-node normal form

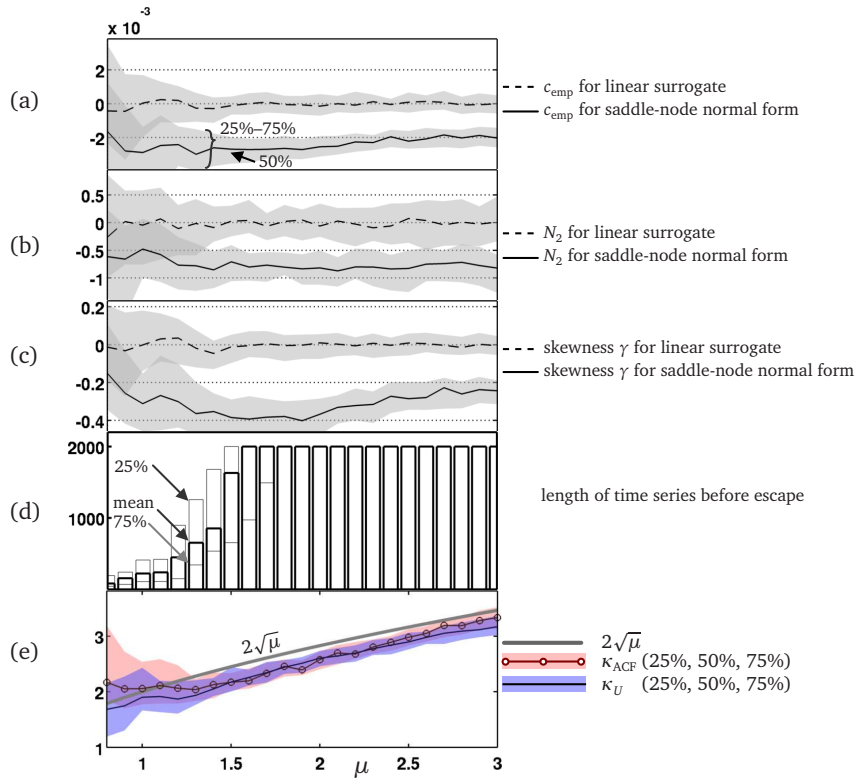


Figure 7: Quartiles of lengths (d), and estimates  $c_{\text{emp}}$  (a),  $N_2$  (b), skewness  $\gamma$  (c),  $\kappa_{\text{ACF}}$ , and  $\kappa_U$  (both (e)) for time series generated by the saddle-node normal form (3.1) with noise ( $\varepsilon = 0$ ,  $\mu = 1$ ) and for linear time series generated by (4.2). Without nonlinearity transition would occur at  $\mu = 0$ .

Figure 7 shows how the estimates for nonlinearity behave for the saddle-node normal form with noise, Equation (3.1) with  $\varepsilon = 0$  and  $\sigma = 1$ . Each panel shows the quartiles of the distribution of the estimate for 100 realisations of time series generated by (3.1). Each realisation was run until it reached 2000 data points or  $x = -1$  (indicating escape). Panel (a) shows the estimate  $c_{\text{emp}}$  obtained by linear-least-squares fitting of the approximate stationary Fokker-Planck equation (3.5) to the empirical stationary density  $p_{\text{emp}}$ , panel (b)

shows the quadratic coefficient  $N_2$  obtained from (4.1), and panel (c) shows the skewness  $\gamma$  of  $p_{\text{emp}}$ . For comparison, we generate also 100 linear time series using

$$dx = -2\sqrt{\mu}x dt + dW_t \quad (4.2)$$

and then fit  $c_{\text{emp}}$ ,  $N_2$  and  $\gamma$  for these time series, too. Figure 7 compares the quartiles of the distributions for time series generated by the saddle-node normal form (3.1) and the linear equation (4.2). If there is small overlap in the distribution then the quantity is a good indicator for the presence of a quadratic nonlinear term. We observe that this is the case for all quantities as long as the time series has a length of 2000 (see Figure 7(d) for the distribution of time series lengths). Naturally, the uncertainty, and, hence, the overlap, increases when the time series is shorter due to escape from the potential well to  $-\infty$ . This is the case for smaller  $\mu$  in Figure 7 (see Thompson and Sieber [12] for a quantitative study of escape probabilities for small  $\mu$ ). We also notice that all quantities have a systematic deviation from the theoretical value of the quantity they supposedly estimate:  $N_2$  should be equal to  $-1$  according to (3.1), the real escape rate  $c$  has a much smaller modulus than  $c_{\text{emp}}$  (compare Figure 6(a)), and the skewness  $\gamma$  has theoretically a much larger modulus (compare Figure 6(b)). The deviation for  $N_2$  is relatively small, and is likely caused by the kernel density estimate for  $p_{\text{emp}}$ . The large modulus of  $c_{\text{emp}}$  shows that the linear term  $-\kappa_U \tilde{x}$  is a poor fit for  $-\partial_{\tilde{x}} U(\tilde{x})$ , which makes  $c_{\text{emp}}$  a measure of how much the odd symmetry is broken by  $-\partial_{\tilde{x}} U(\tilde{x})$ , rather than an estimate for the escape rate. We note that  $c_{\text{emp},2}$  as fitted in (4.1) has a realistic modulus (close to zero, not shown in Figure 7). The bias of the skewness  $\gamma$  is mainly due to our restriction to time series that stay inside the potential well. However, the characteristic dip of  $\gamma$  seen in Figure 6(b) is still visible in Figure 7(c). Appendix B addresses the dependence of the empirical skewness on all method parameters.

We also note that the estimates for the linear decay rate,  $\kappa_U$  and  $\kappa_{\text{ACF}}$  are more accurate than the estimates for the nonlinearity but by their nature they cannot distinguish the time series generated by the saddle-node normal form from a linear surrogate.

The summative conclusion from Figure 7 is that one can detect the underlying nonlinearity of the deterministic part of Equation (3.1) by observing  $c_{\text{emp}}$ ,  $N_2$  or the skewness  $\gamma$  if the time series is moderately long. One can expect this to be true in the more general case of a time series generated by a system with deterministic dynamics close to a fold and additive noise. While one can in principle recover the underlying parameter  $\mu$  from the estimates for  $\kappa_U$ ,  $N_2$  and  $\sigma$  (see Thompson and Sieber [12]), the uncertainty in  $N_2$  propagates dramatically into uncertainty for  $\mu$  (one has to divide by  $N_2$ ).

The results of Figure 7 show the stationary case ( $\varepsilon = 0$ ). For slowly drifting time series the same analysis will then have to be applied to parts of the time series in sliding windows. We demonstrate this for two geological time series in Section 5. For time series with rapidly drifting system parameter  $\varepsilon$  the estimates for the softening,  $c_{\text{emp}}$ ,  $N_2$  and the skewness  $\gamma$ , will not give a detectable difference from the corresponding linear time series before tipping. The conclusion one would draw from this absence of detectable softening would be correct: tipping happens when the linear decay rate  $\kappa$  reaches the critical value 0 (or slightly later, see Thompson and Sieber [12]). The same holds if the noise level is low, which is equivalent to rapid drift (see Thompson and Sieber [12] for the transformation). In practice one uses the argument the other way round: if the nonlinear estimates do not give a value significantly different from zero one will conclude that the ratio between noise level and drifting speed is small.

## 5. Studies of Geological Time Series

We now apply our nonlinear investigations to two ancient climate tippings. The first is the major transition marking the end of the latest ice age which occurred about 17,000 years before the present (yrs BP). The data used for this is a temperature reconstruction from the Vostok ice core deuterium record (Petit et al. [22]). The second more recent tipping is the ending of the Younger Dryas using the grayscale from the Cariaco basin sediments in Venezuela (Hughen [23]). This Younger Dryas event was a curious cooling, about 11,500 years ago, just as the Earth was warming up after the last ice age. It ended in a dramatic tipping point, when the Arctic warmed by  $7^\circ\text{C}$  in 50 years. This sudden ending has been related (Houghton [24]) to a *switching-on* of the global oceanic thermohaline circulation (THC). This switch-on is known from extensive theoretical studies (Dijkstra and Weijer [25]) to be at a saddle-node fold arising as a perturbation of a sub-critical pitchfork (see, for example, Rahmstorf [26] and Thompson and Sieber [12]).

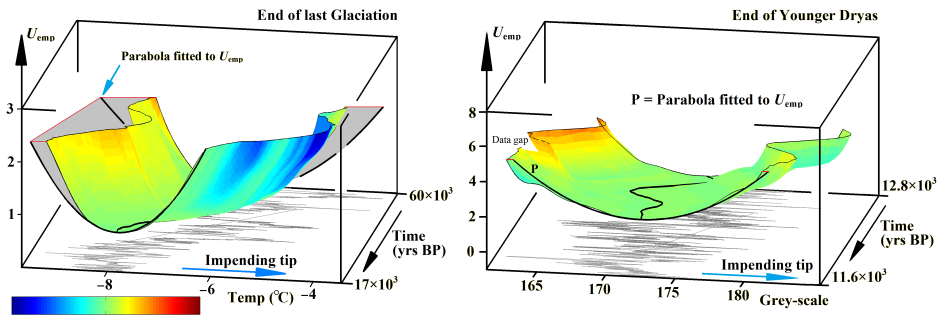


Figure 8: Two predicted potential energy surfaces for (a) the end of the last glaciation, using the Vostok ice-core record, and (b) the end of the Younger Dryas event when the Arctic warmed by  $7^\circ\text{C}$  in 50 years, using the grey-scale of basin sediment in Cariaco, Venezuela. The colour shows the deviation from a (time-dependent) parabola fitted to  $U_{\text{emp}}(x,t)$ , three of which are illustrated above. Red signifies a positive deviation, blue a negative deviation. Time is given in years before the present (BP). The plotted potential was obtained via  $U_{\text{emp}} = -\sigma_{\text{emp}}^2 \log(p_{\text{emp}})/2$  and (3.7) and sliding windows of half the length of the record.

Under linear time-series analysis of the directly preceding data, neither of these two events shows a strong trend in the stability propagator (see Figure 9). Meanwhile sample estimates of the underlying potential functions are shown in Figure 8.

In Figure 8(a), for the end of the last glaciation, we see clearly that the well is softening (falling beneath the parabolic fit) on the high-temperature end while hardening (rising above the parabola) on the low-temperature end. So there is a strong nonlinear signal that a jump to higher temperatures may be pending (as indeed it was). We show that this signal is statistically significant in our full analysis in Figure 9. By comparison, the study of the Younger Dryas, illustrated in Figure 8(b), provided no significant nonlinear conclusions.

For Figure 9 we detrended both time series using Gaussian filtering (see Figures 9(c) and (d)) and estimated the linear indicators  $\kappa_{\text{ACF}}$  and  $\kappa_U$ . We observe weak to non-existent trends, leaving the evidence inconclusive at the linear level (grey and black curves Figure 9(e) and (f)). The ratio between  $\kappa_{\text{ACF}}$  and  $\kappa_U$ , both shown in Figure 9(e) and (f), gives an estimate of the variance  $\sigma^2$  of the noise input. The unit for  $\kappa_{\text{ACF}}$  is *units of  $x$  per time step* (so, for example for Fig. 9 (e)  $^\circ\text{C}/\Delta t$ ), and, correspondingly, the unit for  $\kappa_U$  is  $\Delta t/^\circ\text{C}$ .

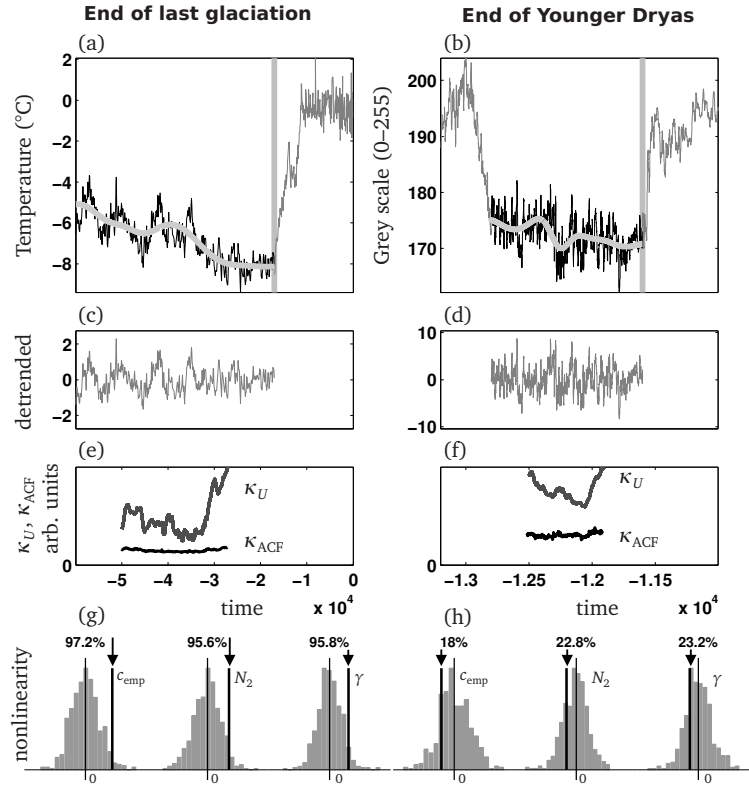


Figure 9: Linear and nonlinear coefficients for two paleo-climate time series. (a,c,e,g): End of last glaciation (snapshot of data from Petit *et al.* [22]); (b,d,f,h): End of Younger Dryas (snapshot of data from Hugen [23]). Only the black part of the data in panels (a) and (b) was used in the analysis. Panel (c) and (d) show the time series after detrending. Panel (e) and (f) show the linear indicators  $\kappa_{ACF}$  and  $\kappa_U$ . Panel (g) and (h) show the means of the estimates  $c_{emp}$ ,  $N_2$  and the skewness  $\gamma$ , compared to zero (thin vertical black line) and a histogram of estimates sampled from 500 random linear time series generated by the linear process (4.2) with  $-2\sqrt{\mu} = \kappa_{ACF}$ . Sliding window length  $w = N/2$ .

(unit of  $\kappa_{ACF}$  divided by unit for  $\sigma^2$ ). Both units are arbitrary such that we do not indicate the scale of the y-axis in Figure 9(e,f).

[Lenton *et al.* this volume] discuss the linear analysis of both time series in greater detail. Their time series *Vostok* corresponds to Figure 9(a), their time series *Cariaco* corresponds to Figure 9(b). (Note that [Lenton *et al.* this volume] use the deuterium proxy directly whereas Figure 9(a) shows the temperature reconstruction.) Specifically, Figures 2 and 4, and Figures 9 and A1 in [Lenton *et al.* this volume] discuss how the results depend on method parameters such as detrending bandwidth and sliding window. Their analysis finds that the presence of a significant trend depends strongly on both parameters. See [Lenton *et al.* this volume], Table 1 (rows *Vostok* and *Cariaco*), for a summary of the evidence at the linear level.

At the nonlinear level, the time series for the end of the last glaciation has a strong quadratic nonlinearity in  $\partial_x U(\tilde{x})$  where the potential well  $U$  is presumed to govern the deterministic part of the dynamics (Figure 9(g)). The indicators  $c_{emp}$ ,  $N_2$  and the skewness  $\gamma$  in Figure 9(g) are all far removed from what can be expected by chance in a linear time series. The histogram in the background of Figure 9(g,h) has been sampled from 500



random linear time series generated by the linear process given in Equation (4.2) with  $-2\sqrt{\mu} = \kappa_{\text{ACF}}$  where  $\kappa_{\text{ACF}}$  is taken from the estimate shown in Figure 9(e). The percentages at the top of Figure 9(g,h) express how far in the tail of the histogram the mean of the quantity extracted from the geological time series is. We warn that the three quantities  $\gamma$ ,  $N_2$  and  $c_{\text{emp}}$  are not really three independent indicators as they all depend on the same estimate of the empirical density  $p_{\text{emp}}$ .

Visual inspection of the well shape in Figure 8 confirms that the well is softening (bending downward) on the high-temperature end and hardening (bending upward) on the low-temperature end. So, at the nonlinear level the time series data close to the bottom of the well gives already evidence for a propensity to escape toward larger temperatures. The time series for the end of the Younger Dryas does not show evidence for strong nonlinearity of the underlying dynamics at the second-order level (Figure 9(h)). The values for  $c_{\text{emp}}$ ,  $N_2$  and the skewness  $\gamma$  can all be explained by randomness as the histograms of estimates for the 500 linear time series (as the histograms in Figure 9(h) show). Note that we did not include the previous transition (visible at the left end in Figure 9(b)) into our analysis as our aim was to infer the nonlinearity exclusively from the data near the tipping equilibrium.

## 6. Conclusion

The main message of the paper is that the linear analysis investigated by [Lenton *et al.* this volume] on its own, even if all estimates are accurate, does not contain all information necessary to estimate the probability of tipping over time. The result of the linear analysis is an estimate of the decay rate  $\kappa$ , which is taken relative to the time spacing of the measurements. Even if this decay rate has an identifiable trend to zero the probability of tipping over time is determined by the dominant nonlinear coefficient  $N_2$  in conjunction with  $\kappa$  and the noise level.

We found that the dominant nonlinear coefficient  $N_2$  is much harder to estimate accurately, so we also looked at proxies that indicate nonlinearity such as the skewness  $\gamma$  (as proposed by Guttal and Jayaprakash [19]) or the coefficient  $c_{\text{emp}}$  from Equation (3.5) (which is also a scalar signed measure for deviation from linear theory). Our study of the saddle-node normal form suggests that both proxies,  $\gamma$  and  $c_{\text{emp}}$ , give values that are easier to distinguish from random chance than the estimate for  $N_2$  itself. A significantly non-zero value of either of these proxies indicates that a quadratic nonlinear term is present and gives its sign.

However, we found the uncertainty for moderately long time series ( $N = 2000$ ) too large to translate the proxy back into a quantitatively reliable estimate for the normal form parameter (this would be necessary to read off the probability for tipping from the tables in Thompson and Sieber [12]). We also found that the nonlinear proxies do not have discernible trends when one approaches tipping, so it makes sense only to extract their overall mean from the time series but not the trend.

## References

- [1] IPCC. 2007 *Climate Change 2007, Contribution of Working Groups I-III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, I The Physical Science Basis, II Impacts, Adaptation and Vulnerability, III Mitigation of Climate Change*. Cambridge: Cambridge University Press.

- [2] Lenton, T. M., Held, H., Kriegler, E., Hall J. W., Lucht, W., Rahmstorf, S. & Schellnhuber, H.-J. 2008 Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences of the United States of America*, 105(6), 1786–1793.
- [3] Scheffer, M. 2009 *Critical Transitions in Nature and Society*. Princeton: Princeton University Press.
- [4] Launder, B. & Thompson, J. M. T. editors 2010 *Geo-Engineering Climate Change: Environmental necessity or Pandora's Box?* Cambridge: Cambridge University Press.
- [5] Held H. & Kleinen, T. 2004 Detection of climate system bifurcations by degenerate fingerprinting. *Geophysical Research Letters*, 31(23), L23207.
- [6] Livina, V. N. & Lenton, T. M. 2007 A modified method for detecting incipient bifurcations in a dynamical system. *Geophysical Research Letters*, 34(3), L03712.
- [7] Dakos, V., Scheffer, M., van Nes, E. H., Brovkin, V., Petoukhov, V., Held, H. 2008 Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38), 14308–14312.
- [8] Lenton, T. M., Livina, V. N., Dakos, V., van Nes, E. H. & Scheffer, M. 2011 Early warning of climate tipping points : comparing methods to improve robustness. *Phil. Trans. Roy. Soc. A*, this volume.
- [9] Thompson, J. M. T., Stewart, H. B. & Ueda, Y. 1994 Safe, explosive, and dangerous bifurcations in dissipative dynamical systems. *Physical Review E*, 49(2), 1019–1027.
- [10] Thompson J. M. T. & Stewart, H. B. 2002 *Nonlinear dynamics and chaos*, 2nd edn. Chichester: Wiley.
- [11] Thompson, J. M. T. & Sieber, J. 2010 Predicting Climate Tipping Points. In *Geo-Engineering Climate Change: Environmental necessity or Pandora's Box?*, chapter 3, pp 50–83, Cambridge: Cambridge University Press.
- [12] Thompson, J. M. T. & Sieber, J. 2011 Climate tipping as a noisy bifurcation: a predictive technique. *IMA Journal of Applied Mathematics*, 76(1), 27–46.
- [13] Ashwin, P., Wicczorek, S., Vitolo, R. & Cox, P. 2011 Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. *Phil. Trans. Roy. Soc. A*, this volume.
- [14] Botev, Z. I., Grotowski, J. F. & Kroese, D. P. 2010 Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5), 2916–2957.
- [15] Livina, V. N., Kwasniok, F., Lohmann, G., Kantelhardt, J. & Lenton, T. M. 2011 Changing climate states and stability: from pliocene to present. *Climate Dynamics*, 37, 2437–2453.
- [16] Ditlevsen, P. D. & Johnsen, S. J. 2010 Tipping points: Early warning and wishful thinking. *Geophysical Research Letters*, 37(19), L19703.

- [17] Kleinen, T., Held, H. & Petschel-Held, G. 2003 The potential role of spectral properties in detecting thresholds in the Earth system: application to the thermohaline circulation. *Ocean Dynamics*, 53(2), 53–63.
- [18] Biggs, R., Carpenter, S. R. & Brock, W. A. 2009 Turning back from the brink: detecting an impending regime shift in time to avert it. *Proceedings of the National Academy of Sciences of the United States of America*, 106(3), 826–831.
- [19] Guttal, V. & Jayaprakash, C. 2008 Changing skewness: an early warning signal of regime shifts in ecosystems. *Ecology Letters*, 11(5), 450–60.
- [20] Guttal, V. & Jayaprakash, C. 2008 Spatial variance and spatial skewness: leading indicators of regime shifts in spatial ecological systems. *Theoretical Ecology*, 2(1), 3–12.
- [21] Beaulieu, C., Chen, J. & Sarmiento, J. L. 2011 Change point analysis as a tool to describe past climate variations. *Phil. Trans. Roy. Soc. A*, this volume.
- [22] Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J.-M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V. M., Legrand, M., Lipenkov, V. Y., Lorius, C., Pépin, L., Ritz, C., Saltzman, E. & Stievenard, M. 1999 Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*, 399(6735), 429–436.
- [23] Hughen, K. A. 2000 Synchronous Radiocarbon and Climate Shifts During the Last Deglaciation. *Science*, 290(5498), 1951–1954.
- [24] Houghton, J. T. 2004 *Global warming: the complete briefing*. Cambridge: Cambridge University Press, Cambridge.
- [25] Dijkstra, H. A. & Weijer, W. 2005 Stability of the global ocean circulation : Basic bifurcation diagrams. *Journal of Physical Oceanography*, 35(6), 933–948.
- [26] Rahmstorf, S. 2000 The Thermohaline Ocean Circulation: A System with Dangerous Thresholds? *Climatic Change*, 46(3), 247–256.

## Appendix A. Sensitivity analysis for estimates of nonlinear quantities in palaeoclimate records

Figure 10 shows how the results in Figure 9(g,h) depend on our method parameters. Specifically, we vary the length of the sliding window and the bandwidth of the Gaussian kernel when fitting the quasi-equilibrium drift. In all panels the  $x$ -axis is the length of the sliding window relative to the overall length of the time series on a logarithmic scale (overall length is  $N = 525$  for the end of the last glaciation,  $N = 2646$  for the end of the Younger Dryas). For example, the  $x$ -value of  $-1$  corresponds to a window length  $w = N/2$ . The  $y$ -axis is the kernel bandwidth for the Gaussian fitting kernel relative to the overall length of the time series on a logarithmic scale (this enters the function `kde1d` as a resolution parameter). For example, the  $y$ -value of  $-3$  means that the resolution parameter of `kde1d` is  $2^3 = 8$ .

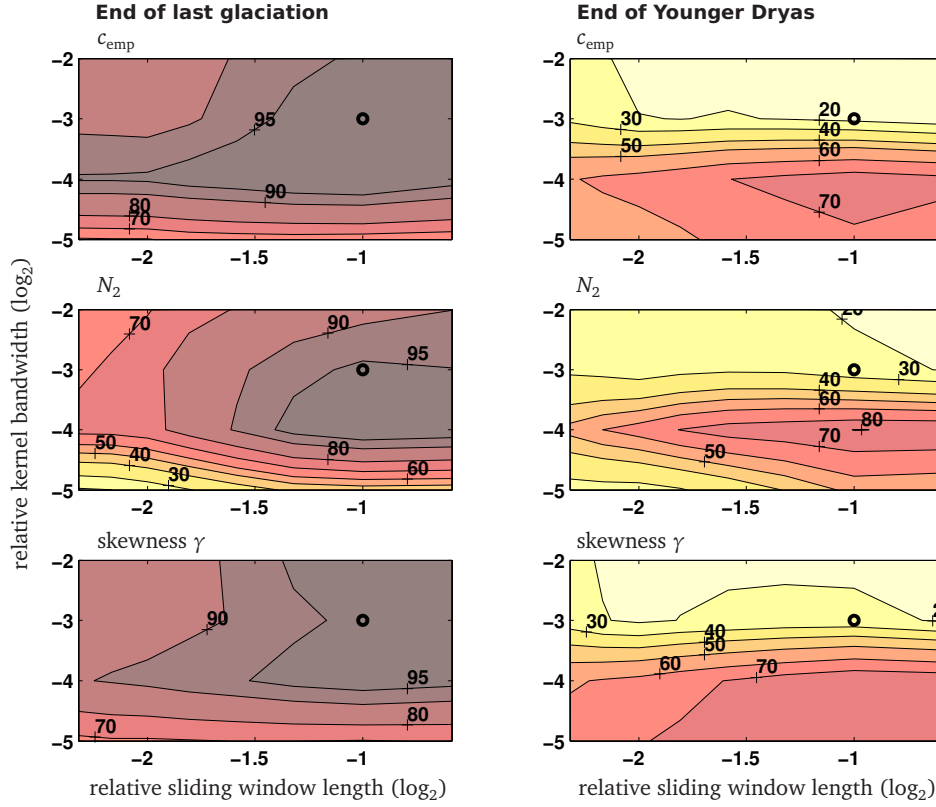


Figure 10: Sensitivity of estimates with respect to the method parameters sliding window length (horizontal) and fitting kernel bandwidth (vertical). The contour levels show every 10th percentile (10%, ..., 90%) and the 95% level (colour online).

For each measured quantity we compare the mean of its value over all sliding windows to the percentiles of the linear surrogate. In this way contour levels close to 0 or 100 correspond to significant deviations from a time series generated by a linear process. The distributions of these time series are shown as gray background in Fig. 9(g,h). The value shown as a black vertical line in Figure 9(g,h) was obtained using the method parameters highlighted by a black circle in Figure 10.

Fig. 10 gives evidence that the nonlinear features of the time series for the end of the last glaciation are robust as long as the detrending remains coarse grained. The end of the Younger Dryas does not show any nonlinear feature that is significantly different from the surrogates for any choice of method parameters.

## Appendix B. Dependence of skewness on distance to tipping point

The observations in Figure 6(b) and Figure 7(c) appear to contradict the statement by Guttal and Jayaprakash [19] that skewness increases as one approaches tipping. Furthermore, there is a quantitative difference between Figure 6(b) and Figure 7(c), even though they show the same qualitative feature: non-monotonicity of the skewness in the control parameter  $\mu$ . Note that Guttal and Jayaprakash [19] studied the same type of tipping as Figure 6, however, not in the fold normal form but in a particular ecological model. In order to inves-

tigate this apparent contradiction we look at how the skewness depends on two parameters: the control parameter  $\mu$  and the value at which we consider a realisation as escaped. More precisely, both figures, 6(b) and 7(c), study

$$dx = [\mu - x^2] dt + \sigma dW_t, \quad (\text{B } 1)$$

where  $\sigma = 1$  without loss of generality and  $\mu$  is constant. Any realisation of (B 1) escapes to  $-\infty$  in finite time almost surely such that (B 1) does not possess a stationary density. In the calculation for Figure 6 we modified the process (B 1) by re-injecting the realisation at  $+\infty$  whenever it escaped to  $-\infty$ . In this way the process has a stationary density given by the Fokker-Planck equation (3.3) with  $-\partial_x U(x) = \mu - x^2$ . However, this is not accurately describing the density that one is interested in: the probability density of the realisations under the condition that the realisation has not yet ‘escaped’. Clearly, this conditional probability depends on what one means by escaped. So, we introduce the escape boundary  $b$  as an additional parameter. Whenever, a realisation gets smaller than the value  $-\sqrt{\mu} - b$  (that is, it has run away beyond the local maximum of the potential by distance  $b$ ) we stop the integration of (B 1). When one considers the set of all realisations before escape then this also leads to a stationary density. This was done for Figure 7 using  $b = -1 + \sqrt{\mu}$  (such that we count a realisation as escaped whenever it reaches  $-1$ ). The re-injection used for Figure 6 and the choice of  $x = -1$  as escape indicator lead to the systematic difference in skewness  $\gamma$  between the figures 7(c) and 6(b). In practice one encounters only the conditional probability as shown in Figure 7. So, it is important to check how the skewness depends on the somewhat arbitrary choice of the value  $b$  indicating escape. Moreover, the estimate shown in Figure 7(c) suffers from the shortness of the time series for  $\mu$  close to 0 (which is typical in practice).

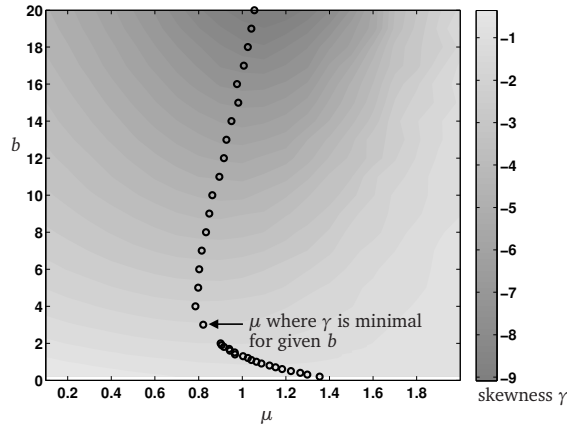


Figure 11: Skewness of probability density of non-escaped realisations

Figure 11 shows the result of a more precise calculation: we simulate (B 1) for a large ensemble of realisations ( $n = 100000$ ). At every time step  $t_k$  we collect the  $n_{\text{esc},k}$  realisations that have reached a value  $x \leq -\sqrt{\mu} - b$ . They count as escaped and have to be discarded. In order to avoid depletion of the ensemble we start  $n_{\text{esc},k}$  new realisations at  $t_k$ . For each newly started realisation we pick the current value of one randomly chosen non-escaped realisation as the initial value. This process reaches a stationary density, which is an approximation of the conditional probability density of  $x$  under the condition that  $x$

has not yet escaped beyond  $-\sqrt{\mu} - b$ . We observe in Figure 11 that the depth of the dip in the skewness depends on  $b$  (indicated by grayscale) but that the location of the dip is uniformly larger than the critical value 0 of the control parameter  $\mu$ .

It is possible that the simulations by Guttal and Jayaprakash [19] stayed consistently to the right of the curve of extreme skewness shown in Figure 11 such that only the increase is observed. Note that in their model the notion of skewness has the opposite sign, so one should consider the modulus of the skewness.